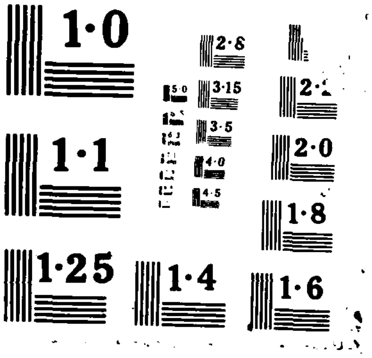
AD-A192 036 LINGUISTIC DEFINITION OF GENERIC MODELS IN COMPUTER 1/1
VISION(U) ROYAL SIGNALS AND RADAR ESTABLISHMENT MALVERN
(ENGLAND) P FRETWELL ET AL. DEC 87 RSRE-MEMO-4110
UNCLASSIFIED DRIC-BR-105253 F/G 12/9 NL


1·0
2·8
5·0
3·15
2·2
3·5
1·1
4·0
2·0
4·5
1·8
1·25
1·4
1·6

(2)

[illegible]

[illegible] No. 4110

[illegible] SIGNALS & RADAR ESTABLISHMENT

[illegible]

[illegible]

[illegible]

PROCUREMENT EXECUTIVE,
MINISTRY OF DEFENCE,
R S R E MALVERN,
WORCS.

ROYAL SIGNALS AND RADAR ESTABLISHMENT

Memorandum 4110

TITLE: LINGUISTIC DEFINITION OF GENERIC MODELS IN COMPUTER VISION

AUTHORS: P Fretwell, P J Goillau

DATE: December 1987

SUMMARY

A method has been developed that can take a human description of an object's spatial appearance and produce a PROLOG representation. The object's appearance is currently in terms of an edge map and the English descriptions are stylised accounts of the salient features and combinations of features found in this representation. At present the translation is performed by hand. However, suggestions are made on how this process can be automated. A prototype translator has been implemented. The PROLOG model is expressed as a hierarchy about the object's appearance, terminating in plausible low-level image primitives. A way is proposed of matching the hierarchy against an image for object recognition in isolation from its background. This reduces the search space of features and feature combinations that the matcher has to consider, so avoiding some of the combinatorial problems when using PROLOG. Extensions using fuzzy logic to deal with uncertain image date and the vagueness of natural language are discussed.

Contents

1 Introduction

There has been great activity in the field of image processing and AI attempting to solve the image interpretation problem. Most results achieved so far have concentrated on simple domains. This simplification has been achieved in a variety of ways but mainly by restricting the class of object to be recognised. Additional techniques need to be developed to deal with the problem of recognition in its generic sense.

The long term objective of this work is to develop tools and techniques for recognition in unconstrained scenes. The main thesis of this work is that unconstrained image interpretation requires novel information structures and processing techniques (Fretwell et al, 1987). It is the purpose of the work to address some of these issues. In particular, it is proposed to develop and demonstrate in principle a suitable knowledge formalism that can be used to represent spatial and non-spatial information for scene interpretation, and secondly to show how this knowledge can be organised and used for recognition. The non-spatial information could include data on the function and context of what is being recognised. Until recently the use of function to represent objects for computer recognition had received little attention in the literature. (See Lowry, 1982, Winston et al, 1983, Adorni et al, 1984, Ingrand et al, 1984, Di Manzo et al, 1985, and more recently Fretwell et al, 1986).

In this paper the following approach is taken. A human-computer interface is used to translate stylised linguistic descriptions of the spatial appearance of objects, generating a form of the description expressed in a logic programming language. The language used in this paper is PROLOG under the POPLOG environment (see Barrett et al, 1985). However, the language FRIL (Fuzzy Relational Inference Language - Baldwin, 1986) is at present being assessed for future use. A taxonomy of subsumption relationships is implicit in the language description. The logic system uses the taxonomy to draw inference about objects from image features extracted from the image by lower level algorithms. In this way the taxonomy is used as a model to match against an image, so providing object recognition.

Using human expertise has been chosen in preference to a machine learning mechanism because current state-of-the-art learning methods cannot cope with the complexity of data presented by a real image.

```
Typical sentences from example English car description:

 "Every car (always) has some wheels".
 "A wheel is_(sort_of)_shaped_like an ellipse".

Predicate calculus representation of meaning:

 all(_1,car(_1) -> exists(_2,wheels(_2) & has(_1,_2)))
 exists(_1,wheel(_1) & exists(_2,ellipse(_2) & is_shaped_like(_1,_2)))

Partial PROLOG database:
 car(X)   :-  has(X,wheels).
 wheel(X) :-  is_shaped_like(X,ellipse).
```

Figure 1: Translation of English to Predicate Calculus to PROLOG

2 Construction of Model

It is first necessary to elicit from human observers their normal English language descriptions of the spatial appearance of the objects concerned. The terms and concepts used in the object descriptions are refined until a bottom level of description is reached. This level represents the interface between the high level object description level and the lower image pixel level. At present it is thought that this intermediate level should comprise two-dimensional features and their interrelationships. Once the object description has been elicited from the human it is transformed by hand from English to PROLOG.

Hand crafting of English car descriptions to PROLOG is appropriate in the short-term, but there are advantages to automatic translation. Naive user descriptions can conveniently be harvested directly into the system, without semantic filtering by the system designer. Knowledge may be added dynamically 'in the field' by extending the hierarchy. In addition to asserting facts and rules the user may ask questions of the system, so giving bidirectional communication.

2.1 Domain Specific Example

A prototype automatic translator has been implemented in POPLOG PROLOG to deal with a limited sub-set of English car descriptions, based on Pereira and Warren's (1980) Definite Clause

Grammar. The syntax of valid sentences is modelled using a Context Free Grammar (CFG) formalism e.g. sentence to noun phrase, verb phrase (see Winograd, 1983). A successful parse involves decomposing a sentence into its bottom-level linguistic primitives (e.g. determiner, noun, verb) matching these terminal symbols to dictionary entries and linking predicate calculus quantifiers, and then reconstructing a predicate calculus representation of the sentence's meaning. Figure 1 gives an example of the translation for some sentences making up elements of a particular car description.

Predicate calculus is a convenient logic formalism that can be used as an intermediate bridge between English and some further representation which is appropriate to the knowledge representation and reasoning mechanism e.g. PROLOG. A series of standard logical manipulations can be used to rewrite a predicate calculus formula into its precise clausal form (Clocksin and Mellish, 1984). The manipulations comprise removing implications, moving negations inwards, Skolemising, moving universal quantifiers outwards, distributing ANDS over ORS, and finally grouping into clauses. The clausal form of predicate calculus is close to a set of PROLOG clauses.

The translator is limited to sentences covered by the syntax and to words held in the dictionary, though both could be extended dynamically with an appropriate user interface. Currently synonyms and non-grammatical input cause the description to be rejected.

3 Object Recognition using the Model

The idea behind the matching strategy is to emulate (however roughly) one possible way in which a human could decide if an object fulfilled the stored descriptive definition. Thus the strategy is essentially top-down or model-driven. The problem of finding suitable object descriptions to match against may be alleviated by using a bottom-up or data-driven partial selection process which isolates likely candidates for matching. This cueing problem is considered elsewhere (Fretwell et al, 1988).

The representation used in the preliminary implementation consists of a PROLOG form of the object's spatial description. The relationship between image, language input and model is shown in Figure 2.

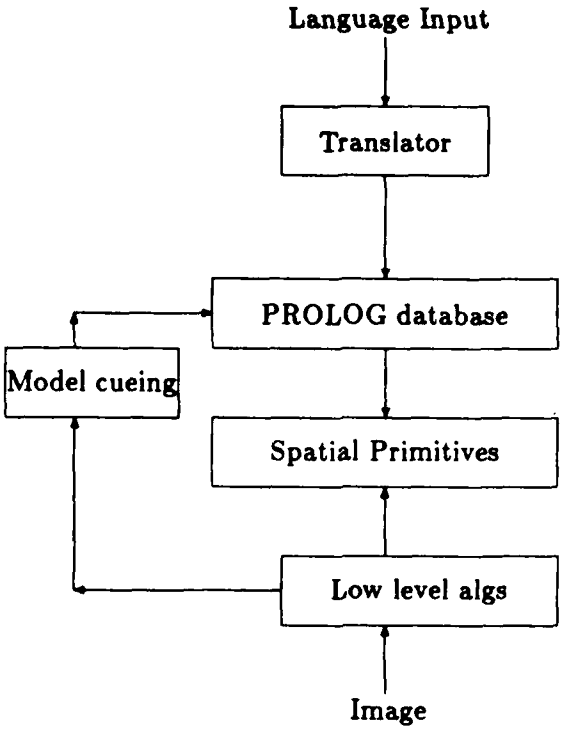


Figure 2: The Matching Paradigm

3.1 Extended Domain Specific Example

In order to see how the recognition paradigm worked with real images the problem of recognising motor cars in natural and man-made environments was chosen, consistent with membership of Alvey consortium MMI/IP 007. A description of the general side view of a car was proposed. The description is as follows:

"A car usually has a roof. A car always has wheels. Wheels are black. They are sort-of shaped like ellipses. Each wheel is placed at a position near the bottom corners of the car. Sometimes only the bottom half of the wheels are visible. A car sometimes has wheel arches. These are concave regions in the body of the car. The wheel arches are roughly semicircular shapes near the bottom of the vehicle. The wheel arches sometimes mask the wheels. A car usually has doors. A door incorporates a window. The door extends from the top part of the car to near the bottom of the car body. A car always has windows. These are closed shapes that have four sides. The windows are near the top of the car body."

Part of the PROLOG representation of the side view of a car is shown in Figure 3. A plausible set of low level spatial primitives that could be used to locate the car within a series of real images

```
car(X):-
        roof(X, ROOF_REGION),
        wheels(X ,WHEELS_REGION),
        wheel_arches(X, WHEEL_ARCHES_REGION),
        windows(X, WINDOWS_REGION),
        car_doors(X, CAR_DOORS_REGION),
        under(CAR_DOORS_REGION, ROOF_REGION),
        under(WHEEL_ARCHES_REGION, ROOF_REGION),
        under(WHEELS_REGION, ROOF_REGION),
        under(WHEEL_ARCHES_REGION, WHEELS_REGION),
        contains(WINDOWS_REGION, CAR_DOORS_REGION).
```

Figure 3: Partial PROLOG Representation of Side View of Car

has been proposed. A subset is shown in Figure 4. For the case of the car the spatial primitive procedures have been written with synthetic values. Figure 5 demonstrates part of the automatic recognition process with the spatial definition of a car acting on a synthetic database. The chain of reasoning has been traced using the "spy" facility of POPLOG PROLOG.

4 Discussion and Future Work

At present the method proposed matches the model against object features. If the background were to become cluttered, or more parts were added to the scene, then the matching mechanism would be affected in the following way. The number of combinations of features necessary to identify the object would become exponentially large as the background and other parts in the image contributed more and more features that did not belong to the object. In general, non-closed world situations provide uncertain and inconsistent data. Thus PROLOG with its hard reasoning is not well fitted to the task of relating the model to the image. Therefore, at present a superset of PROLOG called FRIL (Baldwin, 1986) is being assessed. It is hoped that FRIL with its support logic can accommodate the uncertainty and inconsistency found in real images.

Another major limitation of the proposed method is the potential problem caused by the combinatorial explosion of the search time when using PROLOG (or FRIL) to match the object model to the image. It is clear that some heuristic knowledge must be incorporated into the deduction mechanisms to alleviate the search problems, by using whatever prior knowledge is available to

```
roof(X, ROOF_REGION)
wheels(X, ROOF_REGION)
etc
Some 2D spatial relationships between regions
under(REGION,REGION)
contains(REGION,REGION)
```

Figure 4: Some Plausible Spatial Primitives for Car

generate a best-first search. As has been noted earlier, there are advantages to using non-spatial (e.g. context and function) as well as spatial information in the reasoning process. However, a difficulty of the functional approach appears to be the limited range of objects that can be described adequately by their function. A further difficulty is the problem of implementing suitable functional primitives to interface with the image.

In hand-crafting human object descriptions to PROLOG the system designer applies his own semantic processing which is at present difficult to quantify to extract key concepts and their relationships. With automatic translation of English descriptions there is a problem when mapping the imprecision and vagueness inherent in natural language descriptions (e.g. "sometimes", "mostly") onto the bimodal logic, negation by failure operation of PROLOG. Future work will attempt to map the degree of vagueness onto probabilities in the CFG, using the FRIL package and its support logic mechanism to reason about probabilities. In addition the meaning component may be spread over several sentences comprising a paragraph description. Although the Definite Clause Grammar formalism used in the prototype implementation is powerful and general, there are alternative meaning-based approaches (for example Lexical Functional Grammar, Systemic Grammar, Scripts, Case Frame systems) which are being evaluated for comparison.

5 Conclusions

A method has been developed that can take human analysis of an object's spatial appearance in the form of a stylised description of salient features and combinations of those features, and generate an equivalent PROLOG representation. A paradigm has been proposed by which the PROLOG model may be matched against an object in isolation from its background. The domain specific problem of recognising cars has been discussed by way of example. The car model consists of the salient edges and combinations of those edges. This description is turned into a PROLOG representation of the object by hand crafting. Suggestions have been made on how the language description to

```
** (1) Call : car(_1)?
** (2) Call : roof(_1, _2)
** (2) Exit : roof(_1, [])
** (3) Call : wheels(_1, _3)
** (3) Exit : wheels(_1, [])
** (4) Call : wheel_arches(_1, _4)
** (4) Exit : wheel_arches(_1, [])
** (5) Call : windows(_1, _5)
** (5) Exit : windows(_1, [])
.      .
.      .
.      .
** (1) Exit : car(_1)
```

Figure 5: PROLOG Matching

PROLOG translation could be automated, and a prototype system implemented. A superset of PROLOG called FRIL has been proposed to deal with uncertain image data and the vagueness of natural language.

References

- Adorni, G., Di Manzo, M., Giunchiglia, F., Meson, L., October, 1984, "A conceptual approach to artificial vision", Proc. of the 4th Int. Conf. on Robot Vision and Sensory Controls - RoVisSec 4, London.
- Baldwin, J.F., 1986, "Support logic programming", in Jones, A.I. et al (eds), "Fuzzy sets - Theory and Applications", Proceedings of NATO Advanced Study Institute, Reidel Publishing Co..
- Barrett, R., Ramsey, A., Sloman, A., 1985, "POP-11: A practical language for artificial intelligence", Ellis Horwood, Chichester.
- Clocksin, W.F. and Mellish, C.S., 1984, "Programming in PROLOG" (2nd edn), Springer-Verlag, Berlin.

- Di Manzo, M., Adorni, G., Giunchiglia, F., Ricci, F., 1985, "Building function descriptions : Computer vision", Proc. of the 5th Int. Conf. on Robot Vision and Sensory Controls - RoVisSec, Amsterdam, Netherlands, 29-31 Oct. 1985, pp403-412.
- Fretwell, P., Hearn, D.B., Sleigh, A.C., 1986 "Methods for knowledge representation and inference for image understanding", Oral Presentation at IEE Colloquium on Knowledge-based image processing, Savoy Place, London, 27 January 1986.
- Fretwell, P., Goillau, P.J., Hearn, D.B. and Sleigh, A.C., 1987, "Representation and reasoning using linguistic descriptions of objects", Oral Presentation at BPRA Colloquium on Representing knowledge in vision, Imperial College, London, 7 April 1987.
- Fretwell, P., Bayliss, D.A., Radford, C.J., and Series, R.W., 1988, "Generic cueing in image understanding", this volume.
- Ingrand, F. and Latcombe, J.C., "Functional reasoning for automatic fixture design", November, 1984, CAM-I 13th. Annual meeting and technical conference, Clearwater Beach, Florida.
- Lowry, M.R., 1982, "Reasoning between structure and function", Proceedings of the DARPA image understanding workshops, pages 260-264. Science Applications Inc., McLean, VA., September.
- Pereira, F.C.N. and Warren, D.H.D., 1980, "Definite Clause Grammars for language analysis - a survey of the formalism and a comparison with Augmented Transition Networks", Artificial Intelligence, 13, pp231-278.
- Winograd, T., 1983, "Language as a Cognitive Process. Volume I: Syntax", Addison-Wesley, Reading, Massachusetts.
- Winston, P.H., Binford, T.O., Katze, B., and Lowry, M., 1983, "Learning physical descriptions from functional definitions, examples and precedents", M.I.T. AI Memo, January 1983.

DOCUMENT CONTROL SHEET

Overall security classification of sheet ......UNCLASSIFIED...

(As far as possible this sheet should contain only unclassified information. If it is necessary to enter classified information, the box concerned must be marked to indicate the classification eg (R) (C) or (S))

1. DRIC Reference (if known)	2. Originator's Reference MEMO 4110	3. Agency Reference	4. Report Security Classification UNCLASSIFIED

5. Originator's Code (if known) 778400	6. Originator (Corporate Author) Name and Location RSRE, SAINT ANDREWS ROAD, MALVERN, WORCS WR14 3PS
5a. Sponsoring Agency's Code (if known)	6a. Sponsoring Agency (Contract Authority) Name and Location

7. Title

LINGUISTIC DEFINITION OF GENERIC MODELS IN COMPUTER VISION

7a. Title in Foreign Language (in the case of translations)

7b. Presented at (for conference papers) Title, place and date of conference

PATTERN RECOGNITION, 4TH BPRA CONFERENCE, CAMBRIDGE, MARCH 1988

8. Author 1 Surname, initials FRETWELL P	9(a) Author 2 GOILLAU P J	9(b) Authors 3,4...	10. Date 1987	pp. ref. 10

11. Contract Number	12. Period	13. Project	14. Other Reference

15. Distribution statement

Descriptors (or keywords)

continue on separate piece of paper

Abstract A method has been developed that can take a human description of an object's spatial appearance and produce a PROLOG representation. The object's appearance is currently in terms of an edge map and the English descriptions are stylised accounts of the salient features and combinations of features found in this representation. At present the translation is performed by hand. However, suggestions are made on how this process can be automated. A prototype translator has been implemented. The PROLOG model is expressed as a hierarchy about the object's appearance, terminating in plausible low-level image primitives.
A way is proposed of matching the hierarchy against an image for object recotnition in isolation from its background. This reduces the search space of features and feature combinations that the matcher has to consider, so avoiding some of the combinatorial problems when using PROLOG. Extensions using fuzzy logic to deal with uncertain image date and the vagueness of natural language are discussed.

END
DATE
FILMED
5-88
DTIC